Apr 15th, 12:00 AM

The development of the electronic library

Thomas J. Michalak
Carnegie-Mellon University

Peggy Seiden
Carnegie-Mellon University

THE DEVELOPMENT OF THE ELECTRONIC LIBRARY

Thomas J. Michalak, Director, University Libraries
Peggy Seiden, Librarian for Educational Computing
Carnegie-Mellon University
Pittsburgh, Pennsylvania USA[1]

1. What is the Electronic Library?

The developers of the Carnegie-Mellon University Electronic Library have two goals:

- The development of automated systems to support research and teaching in a university committed to application of new technologies in education
- The creation of systems that disseminate information and teach problem-solving skills necessary to find and use information

The Electronic Library expands the present library beyond its physical location by providing access to local resources in a variety of forms, and directing users to an even wider range of resources beyond the local institution. The Electronic Library under development at Carnegie-Mellon University will enable its users to access a broad range of information resources through a sophisticated, user-cordial electronic work station connected to the University's campus network. Although print resources continue to occupy an important role, electronic communication will serve to link the library's user community directly with local network and remote data sources. These sources include full electronic text of documents, databases of statistical and socio-economic information, locally produced research reports, faculty and staff profiles, commercial databases, and various kinds of directories, dictionaries, and software programs.

2. Background: The Campus Network

The CMU Campus Network Model

The campus network described below provides a unique environment in which the many components of the Electronic Library can be assembled into a coherent working utility. This network comprises individual workstations, software, file servers, and local and remote communications facilities. (See figure 1)

By the end of the decade every student and faculty member on the CMU campus will have access to an inexpensive, high function, individual-use computer. The workstations will typically contain these characteristics:

- Hard Disks with 20 megabytes of storage
- 2-3 Megabytes of RAM and virtual memory (uses hard disk as if it were RAM)
- Multiple Windows with pull down menus
- Ability to run several different programs simultaneoulsy
- Sophisticated graphics displays

Computers will be networked together in a configuration that allows each computer to communicate with all others in the network. The workstations will be connected to a file system which will provide a storage service for individual and campus files. This type of network will allow the creator of information to disseminate his work to the rest of the campus.

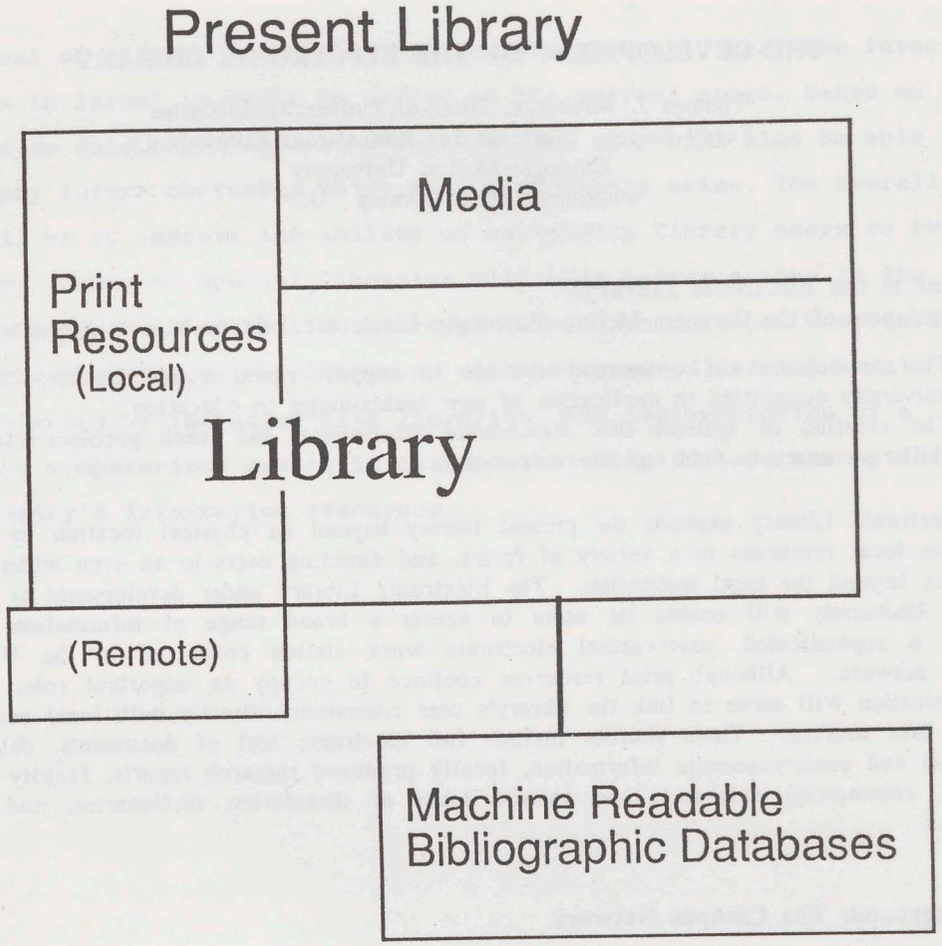

Present Library
Media
Print
Resources
(Local)
Library
(Remote)
Machine Readable
Bibliographic Databases

Figure 1.

Current Network Capabilities

During the past two years the Information Technology Center (ITC) at CMU has made significant progress in developing the campus network. At present no personal computer with the attributes specified earlier is available cheaply, that is, for less than $4,000. Consequently ITC is developing the system software on an interim machine from Sun Microsystems. These machines have the full functionality of the planned workstations.

The ITC has chosen the UNIX operating system, Berkely 4.2, which provides virtual memory, multiple processes, inter-process communication and inter-machine communication. This system has two chief advantages: our applications will not quickly outgrow the hardware, and applications developed on the system are easy to transport to different hardware environments.

System Software:The CMU development system is known as Andrew, named for Andrew Carnegie and Andrew Mellon. Andrew has two components. The first component is a file system, Vice, which will eventually allow 10,000 workstations to be linked together. This network is twice the size of the TWA reservation system, currently the largest network of computers anywhere (5000 terminals). Currently approximately 80 Suns are on the network (Ethernet). There is a file server for every 15 workstations. Software programs and individual files reside on the file server. When a user requests a particular program or file the system can retrieve it from any file server.

The second part of the system software is the user interface, Virtue. At present its major components are a Window Manager and a Base Editor Toolkit. The Window Manager has been operational for over a year. This program executes user commands, arbitrates among competing demands for resources from various applications programs and handles the display of multiple windows on the screen. Thus it allows the PC user to interact with multiple processes, some of which can be running on other machines in the network. It provides a simple, device-independent graphics interface to all applications programs.

The Base Editor toolkit, a software layer added to the window manager, is an interactive text editor that supports standard document preparation styles, creates printed copy on a laser printer, handles virtually all displayed text in the system so that moving text between data-bases, printable documents and conventional tele-type oriented computers is straight-forward.

Additional development of system software is ongoing. The current emphasis is on graphics and tools which can meet the expressed needs of the applications developers.

Applications Software: Fifty Suns have been deployed by ITC to a selected group of applications developers, including the library. The machine described above has the power and flexibility to bring to reality the Electronic Library.

3. The Electronic Library

The Electronic Library Model

Patricia Battin, Vice President and University Librarian at Columbia University describes the workstation as the key that will "provide the electronic scholar the same kind of universal gateway across recorded knowledge as the traditional library provided for printed materials." (See figures 2 and 3)

With appropriate software development the workstation will be capable of:

- Providing access to a wide variety of information sources both on and off campus through a common user-cordial interface

Electronic Library

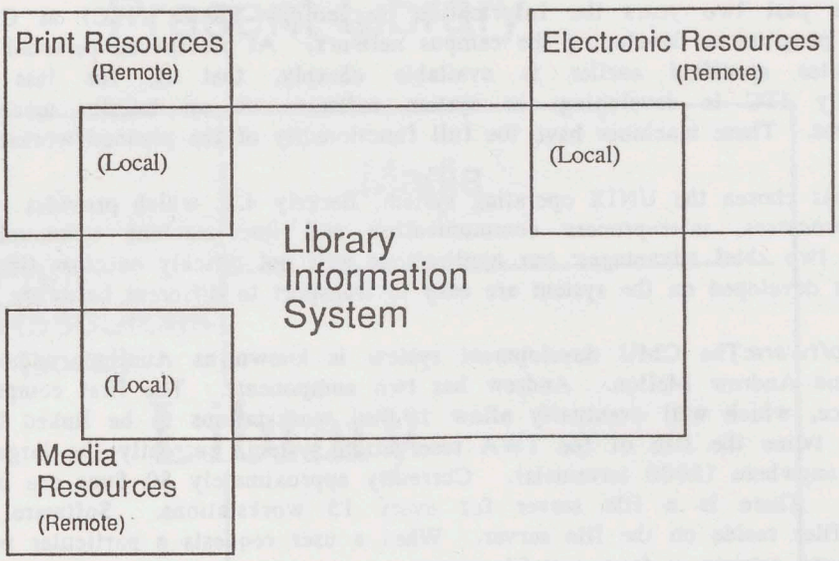


Figure 2.

Library Electronic Workstation

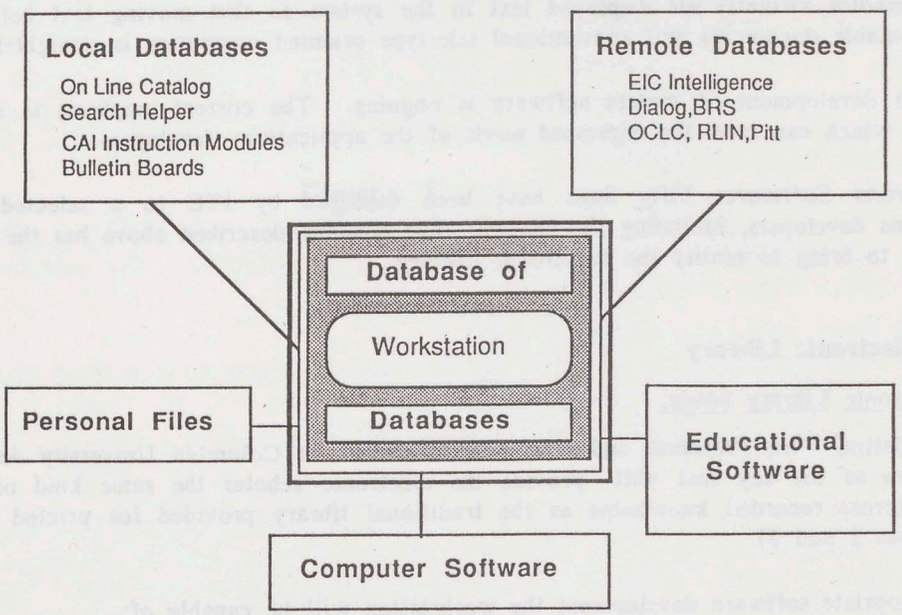


Figure 3.

- Coordinating the use of several information resources simultaneoulsy
- Teaching information seeking skills during the actual process of information retrieval
- Packaging the results of library searches in a format compatible with the text editor
- Evaluating the use of library resources by unobtrusively collecting data

The Real Electronic Library

The ultimate goal of the CMU project is to bring the capabilities of the electronic library to fruition. However, full implementation may be as much as a decade away. Some facets of the electronic library can be achieved in the near future. The remainder of this talk will discuss in detail what the electronic library may look like two or three years down the road and will outline some of the progress we have made toward reaching our intermediate goals and some of the problems which we face.

The major requirements for implementing the Electronic Library are:

- The development of an interface which will provide a common framework for searching a variety of databases based on a model of information retrieval which is comprehensible to the average student.
- The development or acquisition of appropriate information sources.
- Developing methods to provide better access to remote resources, through site licensing and better communications.

(See figure 4).

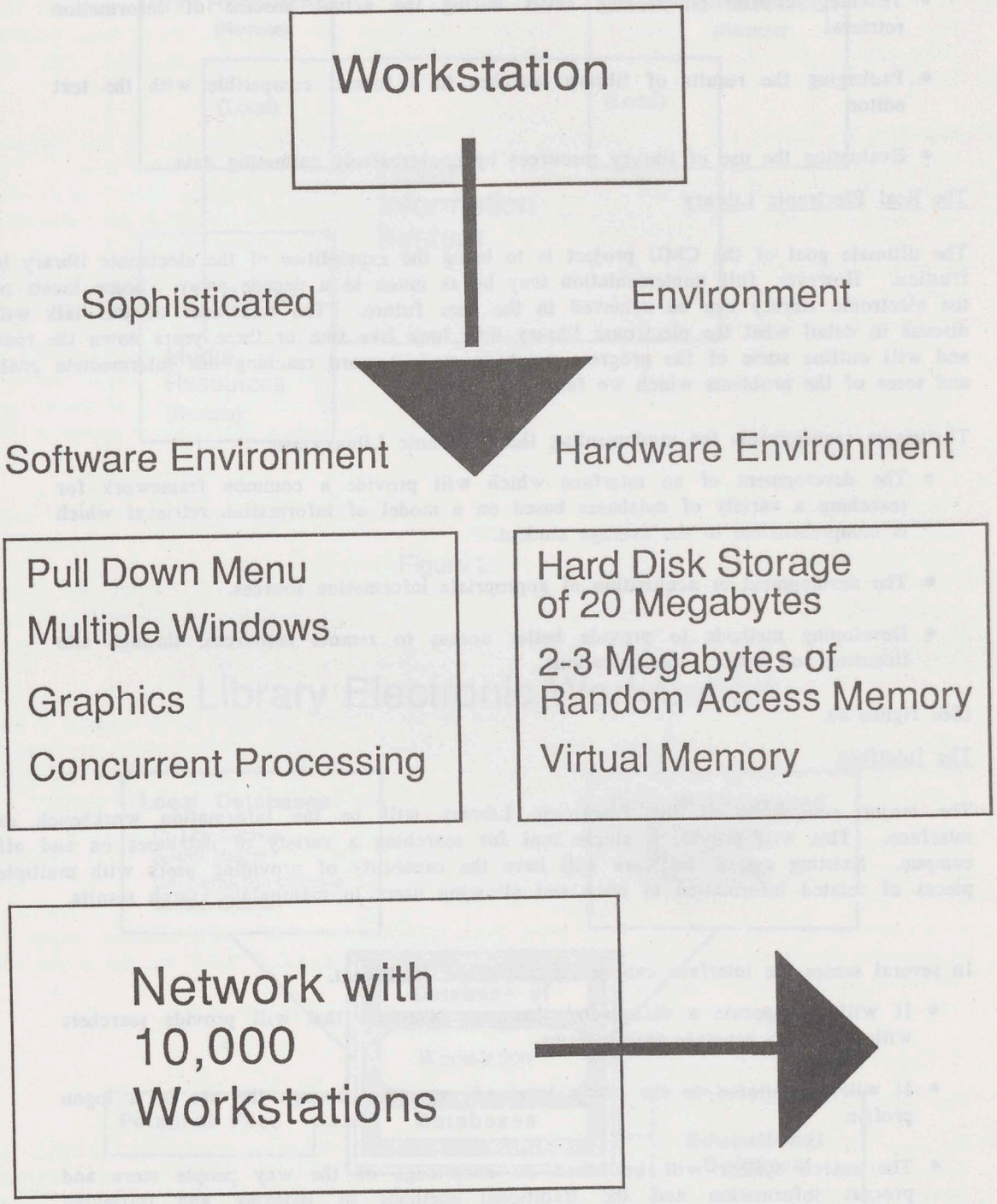
The Interface

The major component of the Electronic Library will be the information workbench or interface. This will provide a single tool for searching a variety of databases on and off campus. Existing system software will have the capability of providing users with multiple pieces of related information at once and allowing users to manipulate search results.

In several senses the interface can be thought of as intelligent.

- It will incorporate a dictionary/ thesaurus structure that will provide searchers with a tool to generate search terms.
- It will be tailored to the user's level of expertise, through the use of a logon profile.
- The search system will be based on knowledge of the way people store and process information and the traditional methods of indexing and retrieving information. The interface will visually model information two retrieval strategies, building blocks and citation pearl growth, that assist the user who does not have a detailed knowledge of various database structures.[2]

The interface software will direct the user's query to the appropriate resources. Two possible


Workstation
Sophisticated Environment
Software Environment
Hardware Environment
Pull Down Menu
Multiple Windows
Graphics
Concurrent Processing
Hard Disk Storage of 20 Megabytes
2-3 Megabytes of Random Access Memory
Virtual Memory
Network with 10,000 Workstations

techniques for accomplishing this are a *database of databases* and parallel searching of several databases. The more promising method is to build a databases of databases, which is an index to databases both on and off campus. This index will be something akin to Dialog's file 411 or BRS's Cross, which provide the user with hit rates for the search term's occurence in each database within Dialog or BRS. A second possible technique is to conduct parallel searches of several databases.

At this first level the user will get some feedback as to the most appropriate sources for the query. The system will attempt to answer the user's query through the "cheapest" path. If any reasonable hits are found by searching the local databases, the system will retrieve and display the search results. However if the system fails, or finds relevant information only for remote sources, or the search results are unsatisfactory, the user can modify the search. He may review the previous search in depth or he may choose online help in the guise of various instructional models, or he may choose to search off-campus sources.

A crucial component of the interface software will be a gateway software package which can access remote systems and translate the user's query into the language of the system being searched. When used in conjunction with the semi-intelligent aspects of the interface the user has the capability of doing more sophisticated and creative searching than the present gateway software allows.

The degree to which a student user will have access to remote databases has not yet been determined. Depending on the communication costs, compared to local storage costs and maintenance, the library may opt to mount many of the most commonly searched databases on the campus network file servers.

Developing Machine Readable Information Resources

The Electronic Library will be able to provide access to diverse types of information, just as the library does today. Part of the work which needs to be done is to make available certain types of information in machine readable form. This involves building databases locally or acquiring databases of this information. The group of resources includes:

- Campus Information such as course offerings, university people database/directories (with faculty publications), bulletin boards on special topics, campus calendar, general university catalogs and handbooks, room availability
- Library Catalog and Journal Holdings
- Enhanced Bibliographic Records for Reference Sources
- Vended Bibliographic Databases (Infotrak [Information Access Corporation], EIC/Intelligence databases, etc.)
- Statistical Databases (Interuniversity Consortium for Political and Social Research data and Census Tapes)
- Catalog of Software, both educational and applications, through which the programs could be identified and directly accessed
- Online Resource Guides -- annotated bibliographies which can give the user some direction in getting started with research in a new discipline. Unlike enhanced reference sources these guides are composed of only the best sources in a particular subject area.

Instructional Modules

Depending on the user's sophistication with the system, type of query, and knowledge of the subject area, the user could go in several directions. In the beginning, the interface will probably not be able to answer all queries directly, but rather the user will have to decide

what is the best way to find the answer he is seeking. To aid the user we are developing a series of instructional modules and providing resource guides online. Later the algorithms used in these modules will be incorporated into the interface. As the interface becomes more sophisticated so will the instructional modules.

Because the interface will be easy to use, there is no real need to develop modules on how to search a particular system. Instruction or help will be needed when the intelligence of the question outstrips the intelligence of the software. This may happen with very specific types of information requests or it may happen when a user is searching a new subject area. The modules will address two common search problems:

- Strategies for subject searching: This would include problems with search vocabulary, or help with searching for information on a topic where one component of the search is ill-defined, (e.g. art in physics).

- Strategies for finding specific types of information: This would include specific programs for searching for commonly asked reference questions, such as finding art reproductions, plays, or demographic data. These programs will categorize a query, find out what a person already knows about a query and match this information to the scope and limitations of various reference sources.

The instructional modules will reside on top of the rest of the search software so that the user can move directly from instruction to actual searches.

Two additional instruction modules are Electronic Resource Guides and the Late Night Librarian. A student might be directed to a set of online resource guides. These guides are annotated bibliographies which can give the user some direction in getting started with research in a new discipline. The user can assemble a personal resource guide from several of these sources. The Late Night Librarian will be an interactive menu-driven program that deals with recurring information requests. These programmable reference interactions will categorize a query, find out what a person already knows about a query, and match this information to the scope and limitations of various reference sources.

Progress to Date

Work towards meeting some of the objectives has already begun in the CMU Libraries through the Online Catalog Project, the experimental distribution of interactive information systems (such as Search Helper) and the use of electronic mail for internal and external communication. Following is a of progress report on each of the components of the system.

Interface Development

The library recently received its Sun workstation and work has started on designing an interface. There is already a prototype interface which is based upon the building block model. The design of the interface is based upon studies conducted at CMU and elsewhere on users' search behavior. Our experience has shown that people bring definite notions about how to find information to an electronic information system. Their mental models of these systems affect the way they approach systems and ultimately their success and failure in the use. Empirical data from tests of students searching the online catalog, data which has been corroborated with evidence from student use of a gateway package, Search Helper, seems to suggest that people may hold a number of different models of an information retrieval system. In developing the resources described below we will be continuing with user testing in order to gather more data on the models and strategies of both expert and novice users searching a variety of different types of systems.

Developing Local Resources

LS/2000 - The Online Catalog: The University Libraries' online catalog has been available to the public since Spring 1984. The system, LS/2000 was developed at the National Library of Medicine and is currently being vended by OCLC, Inc. As a beta test site we have had the opportunity to help in the software development. Enhancements are forthcoming which will provide users with access to journal holdings and easier subject searching. The library has also developed an information function as part of the catalog which will provide online information about the library, its policies and major reference resources.

Campus Information: The campus is making available many of its publications in machine readable form. Some of these, such as bulletin boards and personnel directories are already available on the workstations.

Electronic Resource Guides: The library already has over 30 resource guides available in machine readable form. Access to these will be provided both through the catalog and through a microcomputer search system. The latter more closely approximates the way that such searching will be done on the electronic library. Testing with this software should provide data on the usefulness of such guides in general and on the usefulness of enhanced records for bibliographic searching.

Educational Software Library and Database: A prototype database of educational software has been running since June 1985. Software has already been developed on the workstations which allow a user to search a catalog of programs and then run any programs from that catalog. This software may be adapted to the information retrieval system that will form the basis of the software catalog.

New Computer Hardware: The Libraries plan to use the University's IBM 3038 mainframe computer to provide enhanced searching of the library online catalog and locally mounted databases. IBM donated this machine as part of the continuing IBM-CMU development project. The increased power and storage available with the 3083 will allow the Libraries to test Electronic Library components before the campus network is fully operational.

Access to Remote Databases

Search Helper: For the past 18 months the library has been providing end user searching of databases produced by Information Access Corporation. These databases cover general journal literature, newspapers, business, management, legal and computer journal literature. The system is menu driven and uses an implicit Boolean AND. There are a number of limitations with this system not the least of which is the fact that it is only available during the hours when the reference desk is staffed. Furthermore its surface simplicity can lead to false assumptions and frequent failures on the part of the user.

End User Searching via a Local Database: A second project has taken the problem of searching to a much smaller domain. The fine arts librarian has built a database of citations to articles and papers on designing educational facilities using Sci-Mate. Master's level architecture students were given a seminar on the basic principles of searching before searching the database and are performing their own searches on this limited database.

End User Search Project: The library is now seriously studying end user searching of remote databases such as DIALOG. In Fall 1985 the library plans to provide access to such databases on a trial basis through EasyNet and Wilsearch. As we become more involved with the end user search project we will begin analysing the use patterns of various databases in

order to make "collection development" decisions about which databases are worth the cost of maintaining. These results can be compared with our current statistics of commercial database use.

4. Problems to Surmount

What are the major problems which hinder progress towards the model Electronic Library? Although the interface may seem to be the biggest roadblock toward realization of the Electronic Library, the major problems have little to do with what software can be written.

The following is a list of the problems yet to be resolved and the roadblocks to be got round:

- Communications Problems: To provide high speed communications through the local area network, and to provide communications which can optimally transfer information from large databases.
- Legal Problems: Providing access to copyrighted information and private databases.
- Billing Problems: How to handle the costs of searching remote systems.
- Access Problems: Providing workstations for everyone on campus. By Fall, 1986 there will be 300 publicly available workstations and 800 faculty workstations.

5. Conclusion

Yet even once we solve the problems of technical implementation we have to consider the consequences of the Electronic Library. It is inevitable that the Electronic Libary will have a profound effect on the functions and organization of the traditional library, but it may also have an effect on the quality of an individual's research, vis-a-vis their success in searching.

Providing end-user searching of the major bibliographic databases in the controlled environment of the library is reasonable. What happens when that access is provided from the dorm room, office or lab? How can we ensure search success when we are not around to provide assistance? When a researcher visits the traditional library he's more likely to receive human assistance with search problems even if he comes across that assistance serendiptiously. But the Electronic Library bypasses the possibility of human contact.

Therefore it is even more necessary that the library staff work harder to let the public know why the information specialist is still important and necessary. Providing people with access to large amounts of information easily is not the entire task. Users still need to know how to successfully search for information, and manage and evaluate the information which they have retrieved. Only then will the goals of the Electronic Library be fully realized.

[1]The authors thank Nancy Evans for editorial assistance.

[2]Building Block - Outline or tree concept, vs Citation Pearl Growth - Relating things to a known good article